

AD-A147 494

MODEL ROBUST RESPONSE SURFACE DESIGNS: SCALING
TWO LEVEL FACTORIALS(U) WISCONSIN UNIV-MADISON
MATHEMATICS RESEARCH CENTER D M STEINBERG JUL 84

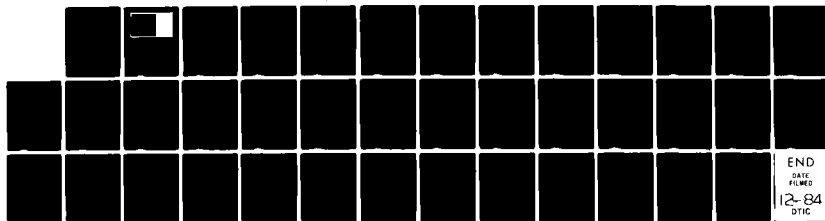
1/1

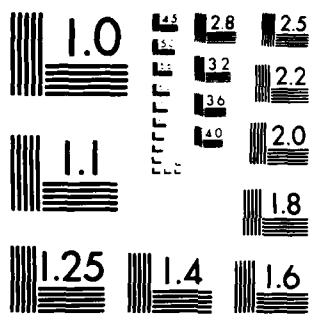
UNCLASSIFIED

MRC ISR-2721 DAAG29-80-C-0041

F/G 12/1

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

3

AD-A147 494

MRC Technical Summary Report #2721

MODEL ROBUST RESPONSE SURFACE DESIGNS:
SCALING TWO-LEVEL FACTORIALS

David M. Steinberg

**Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705**

July 1984

(Received June 19, 1984)

DTIC FILE COPY

Approved for public release
Distribution unlimited

DTIC
NOV 1984

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, DC 20550

84 11 06 053

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

MODEL ROBUST RESPONSE SURFACE DESIGNS:
SCALING TWO-LEVEL FACTORIALS

David M. Steinberg

Technical Summary Report #2721
July 1984

ABSTRACT

Response surface methodology is a useful way to study the relationship between an experimental response variable and a set of continuous explanatory variables. In designing a response surface study, an experimenter must decide how far apart to set the levels of each factor; i.e., how to scale the design. Two conflicting influences must be considered: (i) if the levels are too close together, estimates of the response will have high variance, but (ii) if the levels are too far apart, large bias errors may be introduced. We propose a design criterion based on a Bayesian model that makes explicit assumptions about the possible extent of bias and show that the criterion leads to reasonable choices of scale for 2^{k-p} factorial designs. The choice of scale is found to be insensitive to the prior distributions in the model.

AMS (MOS) Subject Classifications: 62F15, 62K15

Key Words: Response Surface Design; Factorial Experiments; Design Scale; Model Robust Design; Bayesian Models.

Work Unit Number 4 (Statistics and Probability)

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant No. MCS-8210950.

DIU

Accounting	
By	
Distribution	
Availability	
Dist	
A-1	

SIGNIFICANCE AND EXPLANATION

In response surface methodology, carefully designed experiments are used to study the relationship between an experimental response variable and a set of continuous explanatory variables. The experiments are designed to permit estimation of the parameters of a simple graduating function which, it is tentatively assumed, will provide a reasonable approximation to the true response function. These designs usually involve only a few levels of each explanatory variable, so the experimenter must decide how far apart to choose the levels. If the levels are too close together, estimates from the model will have high variance, but if the the levels are too far apart, the graduating function may no longer adequately approximate the true response function, leading to large bias errors. An effective resolution of these conflicting demands must depend on the experimenter's beliefs as to the adequacy of the graduating function. Bayesian statistical methods allow us to formulate a model that includes explicit assumptions about the experimenter's beliefs. ^{The author} We formulate an experimental design criterion based on such a model and ^{study} study the implications of the criterion for scaling two-level factorial experiments, which are often used when the graduating function is a first degree polynomial. ¹¹ We show that the criterion leads to reasonable choices of scale that are not highly sensitive to the experimenter's beliefs.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

MODEL ROBUST RESPONSE SURFACE DESIGNS:
SCALING TWO-LEVEL FACTORIALS

David M. Steinberg

1. INTRODUCTION

Response surface methodology presents a systematic approach to investigate the relationship:

$$E(Y) = g(X_1, \dots, X_k) \quad (1.1)$$

between the expected value of an observed experimental response Y and continuous explanatory variables X_1, \dots, X_k . At the initial stages of a response surface study, it is common to assume that a first-degree polynomial:

$$E(Y) = \beta_0 + \sum_{i=1}^k \beta_i X_i. \quad (1.2)$$

will provide an adequate approximation to the true response function (1.1), at least in an immediate region of interest. (It is assumed in (1.2) that the explanatory variables have been standardized by the experimenter to reflect the region of interest.) Two-level factorial or fractional factorial designs are typically used to estimate (1.2) (see Box, Hunter, and Hunter 1978, Chapter 16).

The problem we consider here is how to select the factor levels in a 2^k -P response surface experiment, i.e., how to scale the design. The choice of scale has important implications for accurate estimation of an unknown response function. If the design points are moved far apart, (1.2) may lead to badly biased estimates of g ; on the other hand, if (1.2) is a good approximation to g and the design points are close to the origin, the estimates will have high variances. Our goal in choosing the scale will be to find

designs that permit accurate estimation of the true response function (1.1) and yet are robust with respect to uncertainty about the functional form of (1.1). Rather than recommending a single "optimal" design, we will suggest a range of reasonably efficient designs. Other criteria might then be used to select a design from this range (see, for example, the list in Box and Draper 1975).

The problem of choosing scale has attracted only scant attention. Rather, most work on experimental design has assumed a fixed design region and then considered how best to allocate the experimental runs within that region. In particular, applications of the theory of optimal design have followed this approach (see, for example, Galil and Kiefer 1977, Pesotchinsky 1978) as have applications of computer-aided design (Mitchell 1974, Mitchell and Bayne 1978, Galil and Kiefer 1980, Welch 1982). Box (1982) criticized the relevance of these studies for response surface experiments, where the experimenter typically has only a vague idea of the limits of the experimental region. In particular, Box (1982) took issue with the conclusion of these studies that many runs should be made at the extreme limits of the design region where an approximate model like (1.2) is most likely to suffer from bias. Our approach reverses the above scheme by considering the allocation to be fixed and then examining how to scale the design. We think that the latter situation is the one actually faced in designing response surface experiments.

Our approach to model robustness is similar to that of Box and

Draper (1959), who studied the effect of model misspecification on the design of response surface experiments. They assumed that the true response function could be written as:

$$g(\mathbf{x}) = \mathbf{f}_1'(\mathbf{x})\beta_1 + \mathbf{f}_2'(\mathbf{x})\beta_2, \quad (1.3)$$

where $\mathbf{f}_1'\beta_1$ corresponds to (1.2) and $\mathbf{f}_2'\beta_2$ represents bias due to quadratic terms. They found that designs which minimize mean squared error for (1.3) are quite similar to those which minimize bias, and showed that minimum bias designs could be found by appropriately scaling the design. Box and Draper (1963) extended their analysis to quadratic approximating functions subject to bias from third degree terms, with similar conclusions. The model form (1.3) introduced by Box and Draper (1959) has proven to be a popular paradigm for investigating model robustness in experimental design. It has reappeared in work by Kussmaul (1969), Stigler (1971), Atkinson (1972), and Jones and Mitchell (1978). All of these papers, however, differed from the original work by Box and Draper in that they assumed a fixed design region and studied the question of allocation.

Any solution to the scaling problem must depend on the experimenter's beliefs as to the ability of (1.2) to approximate g . Our solution is to adopt a Bayesian approach that allows us to make explicit assumptions, in terms of prior distributions, about the adequacy of (1.2) as an approximation to (1.1). Not surprisingly, our recommendations for scaling the design depend on the prior distribution that is used, but a sensitivity analysis

shows that they are quite robust with respect to the prior. In Section 2 we describe a Bayesian analogue of (1.3) and in Section 3 we give some results regarding the precision of estimates based on the model. In Section 4 we propose a criterion for experimental design that is similar to a criterion proposed by O'Hagan (1978) and generalizes a criterion suggested by Wahba (1978). In Section 5 we give results for applying the criterion to 2^{k-p} designs. In Section 6 we examine the implications of the design criterion for 2^{k-p} designs with 4, 8, and 16 factorial runs and one or more center replicates. A discussion of the results is given in Section 7.

2. A BAYESIAN MODEL FOR RESPONSE SURFACES

Suppose we observe experimental data

$$Y_i = g(x_i) + \epsilon_i \quad (2.1)$$

where the ϵ_i are i.i.d. random errors with normal $(0, \sigma^2)$ distributions. Following Box and Draper (1959), suppose the true response function $g(x)$ can be represented as:

$$g(x) = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum_{i=0}^{\infty} \theta_i g_i(x). \quad (2.2)$$

The second summation includes bias due to higher-degree terms and is analogous to the second term in (1.3). We will adopt Young's (1977) suggestion to use orthogonal polynomials for the higher-degree terms rather than simple products of powers. In particular, we will use tensor products of Hermite polynomials, $H_i(t)$, standardized to

have square integral of unity with respect to a normal(0,1) distribution on the real line. Thus (2.2) includes all functions of the form:

$$\prod_{i=1}^k H_j(i)(X_i),$$

where H_j is the one-dimensional Hermite polynomial of degree j .

We will represent the prior belief that a first-degree polynomial is likely to be an adequate approximation to g by assigning uninformative prior distributions to the elements of β but proper priors to the θ_i that constrain these coefficients to be small. Specifically, we assume that:

$$\beta \sim N(0, V) \quad (2.3a)$$

$$\theta_i \sim N(0, \tau \sigma^2 w^{d(i)}) \quad (2.3b)$$

where the θ_i are independently distributed, $d(i)$ is the degree of the corresponding polynomial in (2.2), $w \in [0,1)$ is a parameter that specifies the rate at which higher-degree terms are discounted, and $\tau \in [0, \infty)$ is a measure of the overall extent of bias relative to experimental error. We will make (2.3a) into an uninformative prior by considering limits as $V^{-1} \rightarrow 0$, as in Lindley and Smith (1972). The assumption that the θ_i are independently distributed does not seem unreasonable because of the orthogonality of the regression functions.

3. POSTERIOR VARIANCES

For the Bayesian model (2.2)-(2.3), a natural measure of estimation accuracy is the variance of the posterior distribution of $g(\mathbf{x})$, $\text{Var}\{g(\mathbf{x})/\mathbf{Y}\}$, which we will call estimation variance. The following theorem from Steinberg (1984b) describes the posterior distribution of $g(\mathbf{x})$.

Theorem 1: Let \mathbf{x}_i be a $k \times 1$ vector that lists the factor settings for the i th experimental run. Let

$$\mathbf{f}'(\mathbf{x}) = (1 \ X_1, \dots, X_k)$$

and let \mathbf{X} be the $n \times (k+1)$ matrix whose i th row is $\mathbf{f}'(\mathbf{x}_i)$.

(Primes denote vector or matrix transposes.) Define:

$$R(\mathbf{u}, \mathbf{v}) = \tau \sigma^2 \sum_{i=0}^{\infty} w^{d(i)} g_i(\mathbf{u}) g_i(\mathbf{v}), \quad (3.1)$$

$$\mathbf{r}'(\mathbf{x}) = (R(\mathbf{x}, \mathbf{x}_1), \dots, R(\mathbf{x}, \mathbf{x}_n)) \quad \text{and}$$

$$R_{n \times n} = (R(\mathbf{x}_i, \mathbf{x}_j))_{i,j}.$$

We will assume that \mathbf{X} has full column rank and that the model (2.2)-(2.3) holds with an uninformative prior assigned to β .

Then $g(\mathbf{x})$ has a normal posterior distribution with:

$$\begin{aligned} E\{g(\mathbf{x})/\mathbf{Y}\} &= \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}^{-1}\mathbf{Y} \\ &\quad + \tau \mathbf{r}'(\mathbf{x})\{\mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}^{-1}\}\mathbf{Y} \end{aligned} \quad (3.2)$$

$$\begin{aligned} \text{Var}\{g(\mathbf{x})/\mathbf{Y}\} &= \sigma^2 \{ \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}) + \tau R(\mathbf{x}, \mathbf{x}) \\ &\quad - 2\tau \mathbf{r}'(\mathbf{x})\mathbf{M}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}) \\ &\quad - \tau^2 \mathbf{r}'(\mathbf{x})[\mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}^{-1}]\mathbf{r}(\mathbf{x}) \}. \end{aligned} \quad (3.3)$$

where $\mathbf{M} = \mathbf{I} + \tau \mathbf{R}$.

For the Hermite polynomial expansion of g described in Section 2, we can obtain a closed form solution for (3.1) by slightly modifying (to account for the standardization and the exclusion of the constant and linear terms) Mehler's formula (see Watson 1933):

$$R(u, v) = \frac{\exp\{-(u-v)'(u-v)w^2/2(1-w^2) + u'vw/(1+w)\}}{(1-w^2)^{k/2}} - 1 - wu'v. \quad (3.4)$$

The estimation variance (3.3) is independent of the observation vector Y and is proportional to σ^2 , but is a rather complicated function of the experimental design and the prior parameters τ and w . We can, however, state some general properties:

1. If x is a design point, then $\text{Var}\{g(x)/Y\} < \sigma^2$. This property follows from the fact that conditioning only on the observation made at x would give us an estimation variance of σ^2 ; conditioning on the remaining observations can only decrease the estimation variance. Thus a minimal degree of accuracy can always be assured at any point by taking an observation there. In general, the estimation variance at a point x is decreased when observations are made near x , but may remain almost unchanged if observations are made at distant factor settings. This property is in sharp contrast to the standard conclusion that the variance at x can sometimes be minimized by taking observations far away from x .
2. $\text{Var}\{g(x)/Y\}$ is a monotone increasing function of both τ and w , the prior parameters that state the extent of bias in the model,

and is often approximately linear in τ . Not surprisingly, positing a model with more bias leads to a degradation in the precision of the estimates.

3. Setting either $\tau = 0$ or $w = 0$ eliminates the bias term from the model and (2.2)-(2.3) states that the first degree polynomial is believed to be an exact representation of the response function. The estimation variances in this situation are exactly those that would be obtained from a conventional ordinary least squares analysis of this model. Thus ordinary least squares, by failing to account for the approximate nature of models such as (1.2), can lead to an unduly optimistic assessment of estimation variance.

4. The increase in estimation variance from including bias in the model is especially pronounced outside the range of the data. The Bayesian model agrees with common sense pessimism about the ability to extrapolate from an empirical graduating model.

To illustrate the above comments, and to provide additional insight into the nature of estimation variance for models that explicitly include bias, we consider briefly the estimation variances that result from a 2^3 design under various prior specifications and with different choices of scale. We will assume throughout that $\sigma^2 = 1$.

Figure 1 presents estimation variances for points on one of the coordinate axes when the factors are set at ± 1 for five different priors. The lowest line ($\tau = 0$) gives the ordinary least squares estimation variances. The other priors range from slight bias

($\tau = 1/8$, $w = 0.2$) to moderate bias ($\tau = 1$, $w = 0.4$). The Bayesian estimation variances, although monotone increasing, are relatively flat within the range of the data (i.e., through 1 on the horizontal axis) but increase sharply outside the range of the data.

Figure 2 shows the effect of design scale by graphing the estimation variance functions for a 2^3 design with factors at ± 1 and a 2^3 design with factors at ± 2 . In each case, the prior parameters are $\tau = 1$ and $w = 0.2$. Two slices of the estimation variance function have been plotted for each design, one corresponding to points on a coordinate axis and the other to points on a diagonal of the cube (i.e., points of the form (t,t,t)). In both cases, estimation variance has been plotted against the distance of the point from the origin. It is clear from Figure 2 that using the smaller scale setting provides much better precision at the origin at the expense of high estimation variances outside the sphere of radius $3^{1/2}$ on which the design points are situated. Increasing the design scale permits improved precision across a wider range of values. For points on a coordinate axis, precision is reasonably stable for points within the design cube (i.e., no more than 2 units from the origin). Similar conclusions hold for points along a diagonal. The increase in precision between 2.5 and 3.5 units from the origin corresponds precisely to the design point at 3.46, where the estimation variance must be less than 1. Beyond the design point, estimation variance increases rapidly.

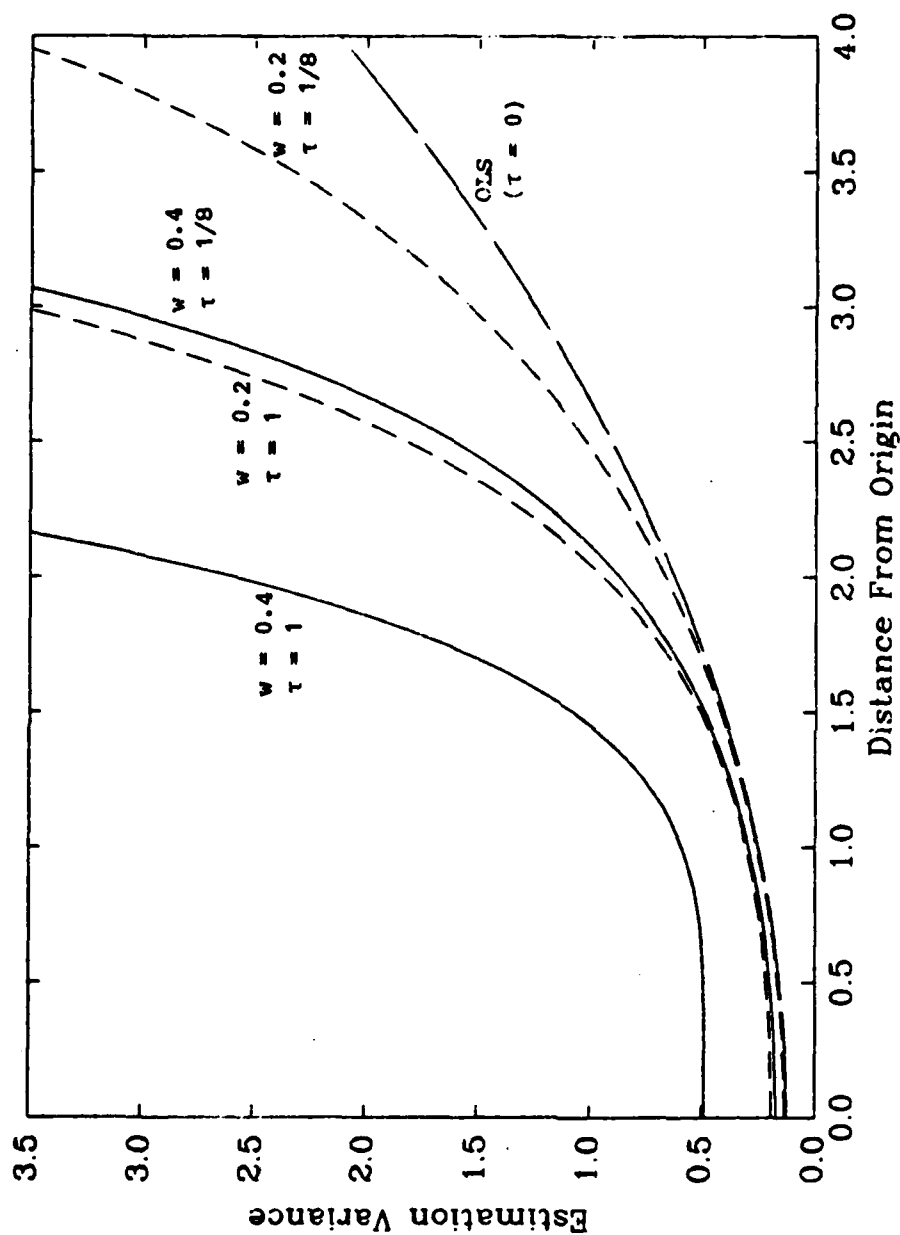


Figure 1: Estimation variance for different choices of the bias parameters in (3.4). Estimation variance is graphed for points on a coordinate axis as a function of distance from the origin. The design is a 2^3 factorial with factors set at t_1 and $\sigma^2 = 1$.

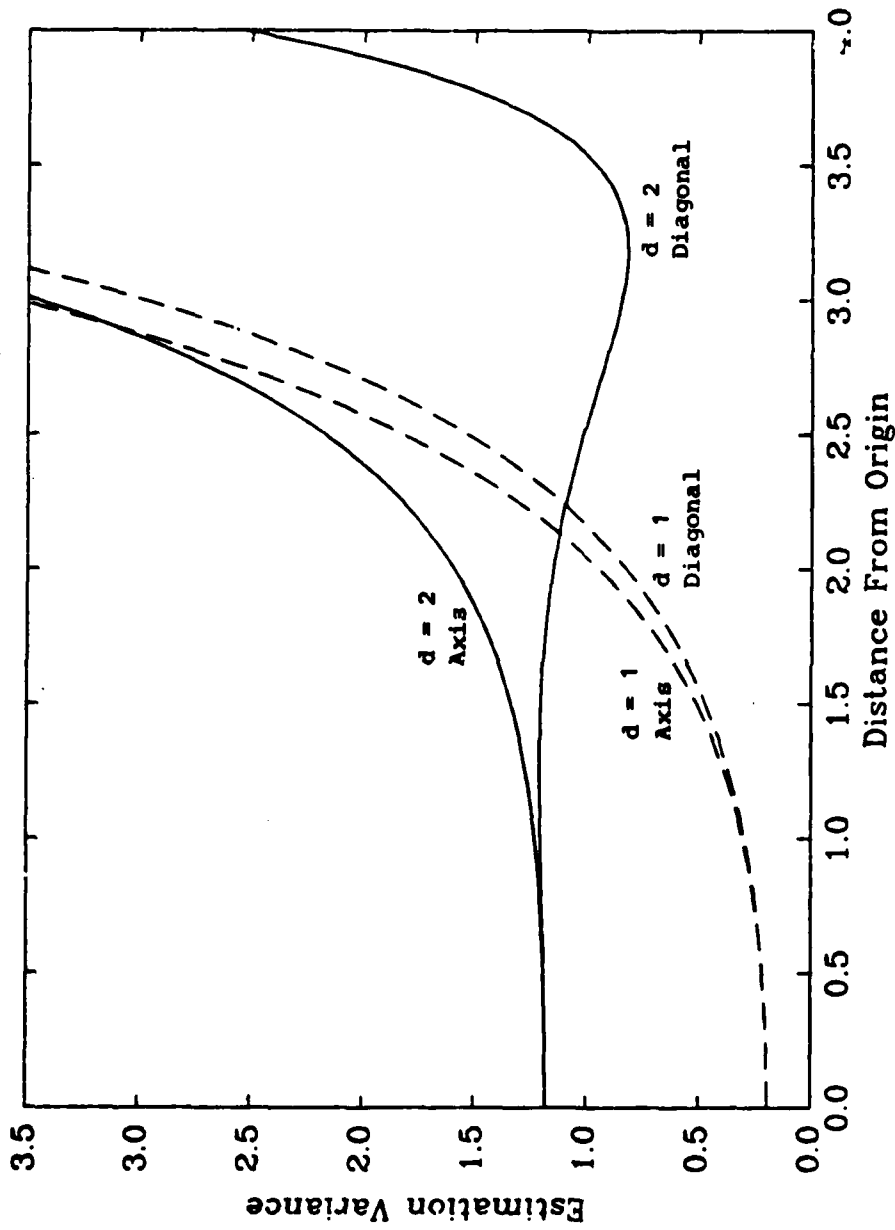


Figure 2: Estimation variance for a 2^3 factorial design with the factors set at ± 1 or ± 2 . For both designs, estimation variance is graphed as a function of distance from the origin for points on a coordinate axis and for points on a diagonal. The bias parameters are $\tau = 1$ and $w = 0.2$. The drop in estimation variance for the wider design between 3 and 3.5 units from the origin on the diagonal corresponds precisely to the design point located 3.46 units from the origin.

4 A BAYESIAN DESIGN CRITERION

The most direct way to use the Bayesian model of Section 2 to compare two experimental designs is to compare their estimation variance functions. Such comparisons are difficult, however, because a design that provides precise estimation in some regions may be uninformative in others. The purpose of a design criterion is to provide a simple means of comparison by giving a numerical summary of the estimation variance function across the entire design region. For our criterion, we define the average weighted estimation variance (AWEV) for an experimental design by:

$$\text{AWEV} = \int_X \text{Var}\{g(\mathbf{x})/\mathbf{Y}\} w(\mathbf{x}) d\mathbf{x}, \quad (4.1)$$

where X denotes the design region and $w(\mathbf{x})$ is a probability density function on X . The p.d.f. $w(\mathbf{x})$ serves as a weight function that reflects the experimenter's interest in different regions of the factor space. Thus AWEV amounts to the expected preposterior loss associated with a (pointwise) squared error loss function and the specified weight function.

The numerical value of the AWEV criterion will, of course, depend on the prior beliefs of the experimenter as to the nature of the response function (i.e., in the case of (2.2)-(2.3), the prior parameters τ and w). This value is of interest in itself, since it summarizes the precision of the estimates that can be made with the model. For comparing designs, however, it is often preferable

to look at relative values of AWEV for fixed prior distributions.

Denoting by Ξ the class of all designs that are under consideration, we define the percent efficiency of a design $D \in \Xi$ by:

$$PE(D) = \left(\min_{E \in \Xi} AWEV(E) / AWEV(D) \right) (100\%) \quad (4.2)$$

Percent efficiency enables us to study questions of robustness with respect to the prior by comparing the efficacy of a particular design across a range of possible priors. Since we do not believe that many experimenters would be able to state unequivocally a prior for (2.2)-(2.3), it is quite important to know how sensitive the choice of design is to the prior specification.

The AWEV criterion is similar to the Bayesian design criterion proposed by O'Hagan (1978), who was also motivated by the problem of scaling experimental designs. O'Hagan's model, although written in a different form than the model in Section 2, is in fact closely related to it (see Steinberg 1984a) and leads to estimation variances of a similar form. The only real difference between O'Hagan's model and that described in Section 2 is the covariance function that corresponds to our (3.4). A problem with O'Hagan's covariance function is that for designs with four or more points, estimation variances are bounded from above even if two of the points are arbitrarily remote from the region of interest (see Steinberg 1983). O'Hagan's design criterion differs from (4.1) because he did not use estimation variance. Instead, he defined a new estimator of the response function which approximates the

posterior expectation estimate (3.2) by a simple parametric function. He defined his design criterion to be the average (weighted) mean squared error of this estimate of $g(\mathbf{x})$ which, in turn, can be decomposed into a posterior variance term (AWEV) and a posterior squared bias term that results from using the simple parametric estimate instead of the posterior mean.

A design criterion even closer to AWEV was proposed by Wahba (1978) in the discussion of O'Hagan (1978). Wahba's criterion, in the notation used here, is:

$$(1 / \tau \sigma^2) E \int [g(\mathbf{x}) - \hat{g}(\mathbf{x})]^2 w(\mathbf{x}) d\mathbf{x}, \quad (4.3)$$

where $\hat{g}(\mathbf{x})$ is the posterior expectation of $g(\mathbf{x})$ and g has a prior distribution as in (2.2)-(2.3) but with proper priors assigned to all the regression coefficients. Fubini's Theorem then justifies interchanging the expectation and integration in (4.3) and, noting that $E\{\hat{g}(\mathbf{x})\} = EE\{g(\mathbf{x})/\mathbf{Y}\} = g(\mathbf{x})$, we see that (4.3) is proportional to (4.1). The two criteria are not equivalent, however. If we assign improper priors to the coefficients in the graduating polynomial, Wahba's criterion (4.3) becomes undefined because $g(\mathbf{x})$ no longer has a formal probability distribution. The posterior distribution of $g(\mathbf{x})$ does exist, however, (provided $\mathbf{X}'\mathbf{X}$ is non-singular), so that the AWEV criterion can be applied in either instance and might be viewed as a generalization of Wahba's criterion.

Wahba's criterion (4.3) is reminiscent of ideas used in numerical analysis for the evaluation of functional approximation

techniques, in which the "closeness" of an approximation $\hat{g}(x)$ to a function $g(x)$ is measured in terms of a norm, such as the average weighted squared difference used above (see, for example, Conte and De Boor 1980, Chapter 6). Since the norm here is stochastic, some summary measure of its distribution must be used; (4.3) summarizes the distribution via its expected value. Thus the AWEV criterion (4.1) can be justified on numerical analytic, as well as statistical, grounds.

Calculating AWEV

Direct computation of AWEV is likely to be intractable in most situations, but a simple identity greatly facilitates the task. Substituting (3.3) into (4.1) gives AWEV as a sum of terms of the form:

$$\int_X u'(x) A v(x) w(x) dx, \quad (4.4)$$

where $u(x)$ and $v(x)$ are vectors that depend on the estimation site x and A is a matrix that depends on the experimental design but not on x . Recall that if L and M are any two matrices such that the products LM and ML are defined, then $\text{tr}(LM) = \text{tr}(ML)$, where tr denotes the trace of a matrix. Applying this identity and some simple algebra, we can rewrite (4.4) as:

$$\int_X u'(x) A v(x) w(x) dx = \text{tr} \left[A \int_X v(x) u'(x) w(x) dx \right]. \quad (4.5)$$

The integral on the right-hand-side of (4.5) involves the experimental design only through vectors of the form $\mathbf{r}(\mathbf{x})$ and is much more amenable to analysis than the integral in (4.4).

A statistically intuitive expression for AWEV can be written using the above identity and the standard statistical expectation operator. Let \mathbf{T} denote a random vector with probability density function $w(\mathbf{x})$. Then:

$$\begin{aligned} \sigma^{-2}_{\text{AWEV}} = & \text{tr}[(\mathbf{X}'\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{E}\{\mathbf{f}(\mathbf{T})\mathbf{f}'(\mathbf{T})\}] + \tau\mathbf{E}\{\mathbf{R}(\mathbf{T},\mathbf{T})\} \quad (4.6) \\ & - 2\tau\text{tr}[\mathbf{M}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{E}\{\mathbf{f}(\mathbf{T})\mathbf{r}'(\mathbf{T})\}] \\ & - \tau^2\text{tr}[(\mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}^{-1})\mathbf{E}\{\mathbf{r}(\mathbf{T})\mathbf{r}'(\mathbf{T})\}]. \end{aligned}$$

where $\mathbf{M} = \mathbf{I} + \tau\mathbf{R}$ and the expectations are taken with respect to the distribution of \mathbf{T} .

5. AWEV FOR 2^{k-p} DESIGNS

In this section we consider the explicit calculation of the AWEV criterion for two-level factorial and fractional factorial designs. Since AWEV is proportional to the experimental error variance σ^2 , we will assume throughout this section and the remainder of the paper that $\sigma^2 = 1$.

In order to calculate AWEV we must specify a weight function $w(\mathbf{x})$ and we will use a standard multivariate normal density for this purpose:

$$w(\mathbf{x}) = (2\pi)^{-k/2} \exp(-\mathbf{x}'\mathbf{x}/2). \quad (5.1)$$

It is important to point out the assumptions made in adopting (5.1) since they will not be appropriate for every experiment. The choice

of a suitable weight function must depend on the units of measurement for the factors. Use of (5.1) implies that these units have been standardized so that the origin is the center of the immediate region of interest and so that the experimenter's interest falls off in a symmetric fashion with increasing distance from the origin. With respect to each individual factor, interest is concentrated on settings between -1 and +1 and is negligible for settings below -3 or above +3. We assume, then, that the factors themselves have been scaled by the experimenter to reflect his region of interest. A crucial point is that we regard this scaling as distinct from the question of design scaling: that the experimenter can standardize the units to match his region of interest does not answer the question of where to place the design points.

Having stated a weight function, we must now evaluate the integrals in (4.6). We first state a general result and then apply it to the special case of 2^{k-p} designs.

Lemma: Let $T = (T_1, \dots, T_k)$ be a random vector with a multivariate normal $(0, I)$ distribution. Let

$$f(T) = (1, T_1, \dots, T_k)' \quad \text{and}$$

$$r(T) = (R(T, x_1), \dots, R(T, x_n))',$$

where x_1, \dots, x_n are points in k -dimensional Euclidean space and, for any two points u and v ,

$$R(u, v) = \frac{\exp\{-(u-v)'(u-v)w^2/2(1-w^2) + u'vw/(1+w)\}}{(1-w^2)^{k/2}} - 1 - wu'v.$$

Then the integrals in (4.6) are given by:

$$(i) \quad E\{f(T)f'(T)\} = I_{k+1}.$$

$$(ii) \quad E\{R(T, T)\} = (1-w)^{-k} - 1 - wk$$

$$(iiia) \quad E\{f(T)r'(T)\}_{i,i} = E\{R(T, x_i)\} = 0$$

$$(iiib) \quad E\{f(T)r'(T)\}_{j+1,i} = E\{T_j R(T, x_i)\} = 0$$

$$(iv) \quad E\{r(T)r'(T)\}_{i,j} = E\{R(T, x_i)R(T, x_j)\}$$

$$= \frac{\exp\{-(x_i - x_j)'(x_i - x_j)w^4/2(1-w^4) + x_i'x_jw^2/(1+w^2)\}}{(1-w^4)^{k/2}}$$

$$- 1 - w^2x_i'x_j.$$

$$= R(x_i, x_j; w^2).$$

We can use the results of the lemma to compute (4.6) when the experimental design is a 2^{k-P} (fractional) factorial with the factors set at $\pm d$, so that d is our design scale parameter. It is clear from the lemma that, for any design, the third term of (4.6) is zero:

$$\text{tr}[\mathbf{M}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{E}\{\mathbf{f}(\mathbf{T})\mathbf{f}'(\mathbf{T})\}] = 0.$$

An additional simplification for factorial designs is also helpful. From considerations of symmetry, the matrix whose trace must be computed in the final term of (4.6) must have all its diagonal elements equal to one another. Thus, only the first column of $\mathbf{E}\{\mathbf{r}(\mathbf{T})\mathbf{r}'(\mathbf{T})\}$ need be computed. If the design also includes center replicates, the final term of (4.6) will require computation of two terms, one corresponding to a point on the cube and one to a center point.

To illustrate the above results, consider scaling a 2^3 design with bias parameters set at $\tau = 1$ and $w = 0.4$. We computed AWEV for $d = 0.05$ (0.05) 3.00; the computations were performed using the MATLAB matrix laboratory package (see Moler 1981) and, along with computations for 9 additional settings of τ , required about 12.5 minutes of CPU time on the VAX 780 computer at the Mathematics Research Center at the University of Wisconsin. The percent efficiencies are graphed in Figure 3. The minimal value of AWEV, about 2.721, is achieved at approximately $d = 1.17$. The efficiency remains high for a fairly broad range of designs, but drops off as the design points are brought in too close to the origin or as they are moved too far away, a conclusion consistent with common experimental wisdom. If, on the other hand, we were to assume that no bias were present ($\tau = 0$), AWEV would be a monotone decreasing function of d , reaching a minimum of .125 at $d = \infty$. The design with $d = 1.17$ would have a much lower AWEV

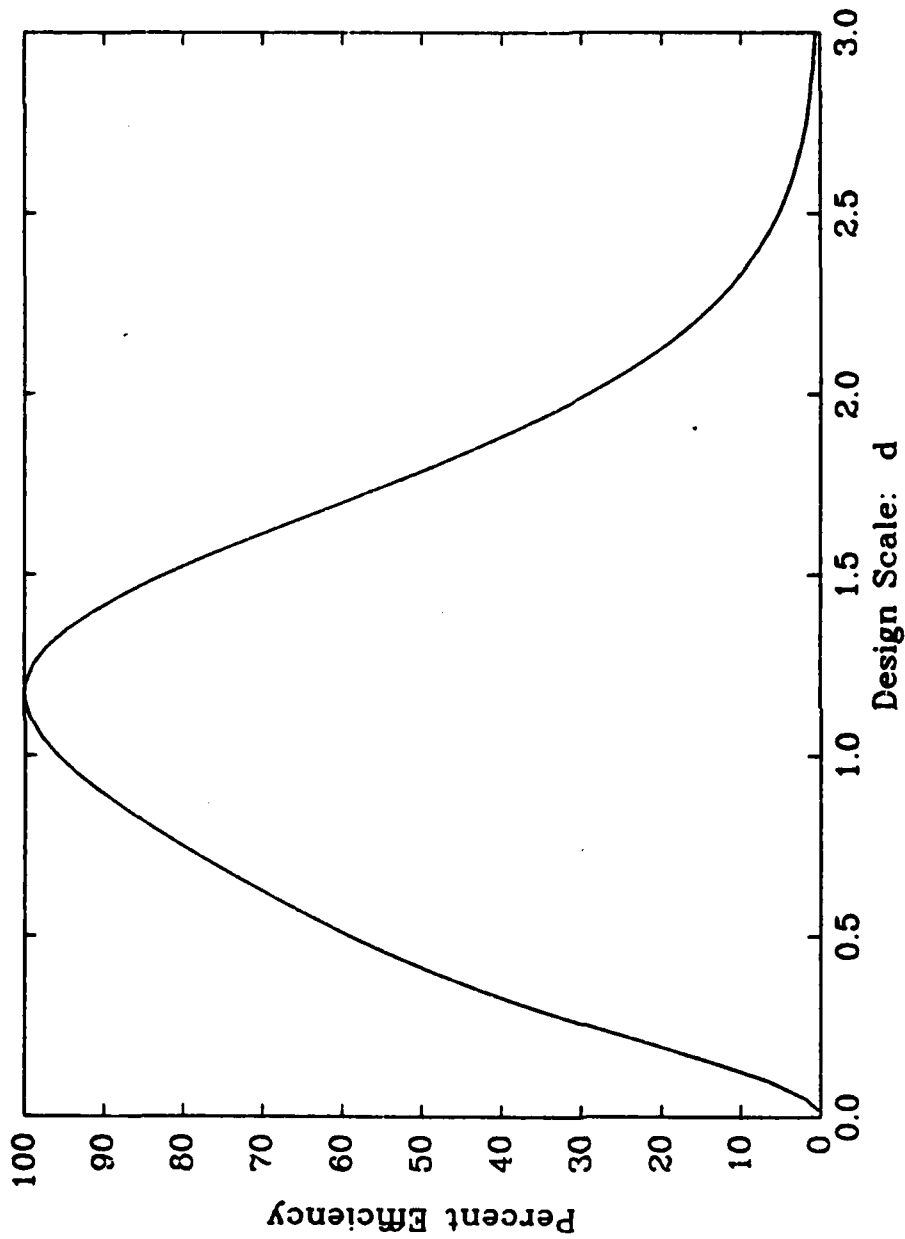


Figure 3: Percent efficiency for a 2^3 factorial design with factors set at t_d , relative to the choice of d that minimizes the AWEV criterion, with the bias parameters set at $\tau = 1$ and $w = 0.4$.

value (.399) but would be only 31.3% efficient.

To study sensitivity with respect to the prior, profiles similar to Figure 3 were generated for a variety of combinations of the parameters τ and w . The values of τ were equally spaced on a log scale and reflect a range of situations from little bias relative to experimental error through substantial bias relative to experimental error; the values of w were $w = 0.1$ (0.1) 0.8. For each parameter combination, as in Figure 3, a range of designs was found to have reasonably high efficiency and efficiency was low for both low and high values of d .

In Figure 4, 90% and 75% efficiency ranges are plotted as a function of τ for $w = 0.2, 0.4, 0.6$, and 0.8 . Highlighting the range of high efficiency designs, rather than just the "optimal" design, allows us to find designs that perform well across a variety of possible experimental conditions; i.e., to find designs that are robust with respect to the prior distribution. It is clear from Figure 4 that an efficient choice of scale for a 2^3 design is quite insensitive to the prior. Choosing d anywhere between 1.1 and 1.6 scales the design efficiently for almost all the parameter combinations considered. Moreover, the slight dependence of the choice of scale on the prior can be easily summarized: the less severe the bias is feared to be, the larger the scale should be. But even for the least severe case of bias here ($\tau = 1/32$, $w = 0.2$), choosing d greater than 3 is quite inefficient. For the most severe case ($\tau = 16$, $w = 0.8$), small choices of d are

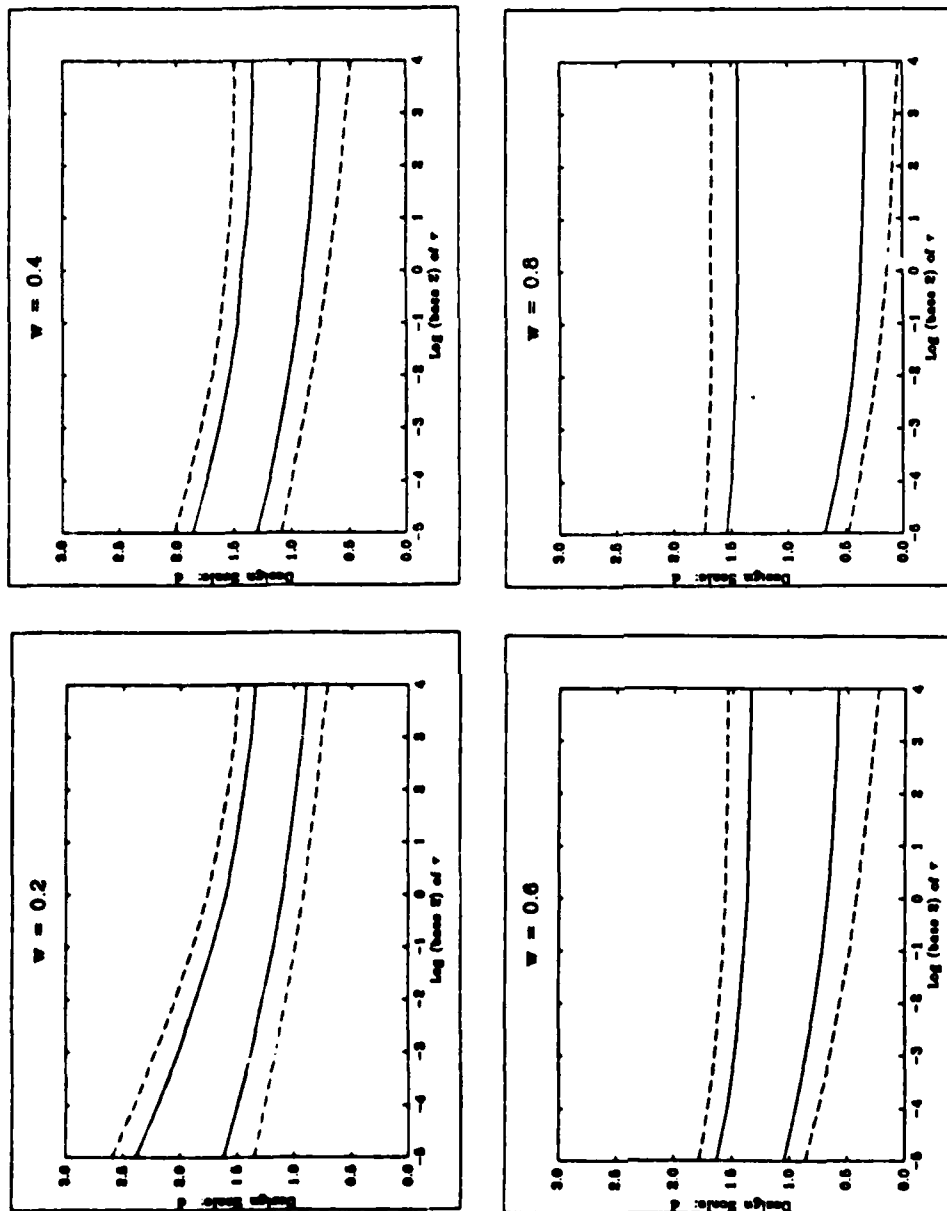


Figure 4: Choices of design scale d that are 90% efficient (within the solid lines) and 75% efficient (within the dashed lines) for a 2^3 factorial design with factors set at τ , for various combinations of the bias parameters w and τ (plotted on a log scale).

relatively efficient, but choosing d as large as 1.6 also retains high efficiency.

It is also informative to examine the AWEV values that an efficient design is able to obtain, since this provides useful information on how precisely the response function can be estimated. The AWEV values, unlike percent efficiency, are quite sensitive to the choice of the prior. When the bias is assumed to be slight, AWEV for the most efficient designs differs only slightly from the AWEV value that those designs would obtain in a model without an explicit bias term. For example, with $\tau = 1/32$ and $w = 0.1$, the minimum AWEV is .205, obtained when $d = 2.57$. For the "no bias" model, this design has $\text{AWEV} = .182$. For choices of scale smaller than 2.57, the effect of bias on AWEV is almost negligible, but for larger choices, bias becomes substantial and AWEV is much larger than in the "no bias" model. When the model includes a large bias term, the results are quite different. When $\tau = 1$ and $w = 0.8$, for example, the minimum AWEV is 125.88 when $d = 0.88$, much greater than in the "no bias" model.

6. SCALING 2^{k-P} FACTORIAL DESIGNS

In this section we describe specific recommendations of the AWEV criterion for 2^{k-P} factorial designs in terms of efficient choices of the design scale parameter d , which specifies the factor settings in units standardized to reflect a multivariate normal $(0, I)$ weight function. We limit our discussion to designs with a maximum of 16 factorial runs (with and without added center replicates) and a maximum of eight factors, which we believe includes many of the most popular two-level fractional factorial designs. Table 1 lists all the designs considered along with the defining contrasts for the fractional factorials. For each design, we computed AWEV and percent efficiency as a function of d for prior specifications in which τ ranged from $1/32$ through 16 and $w = 0.2, 0.4, 0.6, \text{ and } 0.8$.

The results for all of the designs studied follow the same general pattern: a range of scale choices roughly between 0.8 and 1.6 is reasonably efficient for almost all the priors considered but both high and low values of d lead to low efficiency. Percent efficiency is remarkably robust with respect to the prior specification for all the designs studied. Thus precise prior knowledge of the bias parameters is not necessary to obtain an efficient design. Table 1 lists, for each design, the range of scale settings that is at least 75% efficient for all the choices of τ , from $1/32$ through 16, when $w = 0.4$. Even though these bias conditions differ by a factor of 500, there is always a

reasonable range of designs that is efficient across the entire spectrum.

As the extent of bias in the model increases, the range of efficient designs is typically pulled in toward the origin. If, for a fixed value of w , τ is made extremely small, only large choices of scale lead to efficient designs, since the bias term in the model is effectively suppressed. Making τ extremely large, on the other hand, does not have the effect of severely shrinking the range of efficient designs in to the origin. Rather, we find an asymptotic behavior in which increasing τ beyond a certain point ceases to have any effect at all on the percent efficiency of d . Choosing d between 1 and 1.5, for example, rarely results in a design that is less than 75% efficient because it is too far from the origin. Over the range of bias specifications that we studied, the efficient choices of scale are more similar to those for large τ than those for small τ .

The actual values of AWEV are quite sensitive to the prior specification and we might interpret this dependence in two rather different ways. One possible conclusion is that even the most efficient designs are able to provide little information when bias is severe. If so, then the correct decision might be to increase the number of runs, or to use a more flexible graduating function, such as a quadratic, or to limit the region of interest to one in which (1.2) is thought to be a better approximation, or perhaps to scrap the experiment altogether. Alternatively, since (2.2)-(2.3)

Table 1

Listed below are the 2^{k-p} designs studied along with their defining contrasts and the range of scale settings that result in designs of at least 75% efficiency for all values of τ between $1/32$ and 16 when $w = 0.4$.

Design	Contrasts	75% Efficiency Range
4 Runs		
2^2		1.19-1.50
2^{3-1}	I=ABC	1.08-1.33
8 Runs		
2^3		1.09-1.49
2^{4-1}	I=ABCD	0.98-1.37
2^{5-2}	I=ABD=ACE	0.87-1.28
2^{6-3}	I=ABD=ACE=BCF	0.78-1.20
2^{7-4}	I=ABD=ACE=BCF=ABCG	0.71-1.15
16 Runs		
2^4		0.96-1.49
2^{5-1}	I=ABCDE	0.86-1.41
2^{6-2}	I=ABCE=BCDF	0.75-1.34
2^{7-3}	I=ABCE=BCDF=ACDG	0.66-1.28
2^{8-4}	I=ABCE=BCDF=ACDG=ABDH	0.58-1.24

measures bias proportionally to experimental error, the magnitude of the bias term will increase if more precise measurement is able to decrease experimental error. In this case it would be misleading to compare AWEV values for different priors under the assumption that σ^2 is equal, since the smaller value of σ^2 corresponding to the model with greater bias will compensate for the difference in AWEV.

Adding an extra factor It is interesting to examine the effect on AWEV of increasing the number of factors in the design with a fixed number of runs. We would expect the average estimation variance to increase, because we are attempting to study a much larger space with the same number of experiments. For the "no bias" case ($\tau = 0$), AWEV reduces to average weighted least squares variance. Here it is easy to show that the ratio of AWEV for a $2^{(k+1)-(p+1)}$ design with scale parameter d to a 2^{k-p} design with scale parameter d is:

$$(d^2 + k + 1)/(d^2 + k),$$

which implies that there is a noticeable increase only for designs that are too close to the origin. For designs with d large, there is almost no loss of precision at all in adding an extra factor to the experiment. The Bayesian model, however, suggests that when bias is present, there may be a much greater price to pay for adding an extra factor. For example, with $w = 0.4$ and $\tau = 1$, the minimal value of AWEV for a 2^{4-1} design is 6.07 when $d = 0.98$. The corresponding 2^3 design has AWEV of 2.86, less than half as great and much less than the above ratio would indicate. Adding a fifth

factor to the experiment would further increase AWEV to 12.64 and adding a sixth factor would increase AWEV to 23.54. (Despite the increase in AWEV, all these designs are at least 95% efficient.) This pattern is consistent throughout the 8 and 16 run designs. The loss in accuracy from adding an extra factor is most severe when the bias is severe. Thus we recommend limiting the number of factors under consideration when it is feared that there may be substantial bias.

Adding an extra factor also affects the range of efficient scale choices. As can be seen from Table 1, increasing the degree of fractionation tends to pull the range of efficient designs slightly in toward the origin.

Center points It is often recommended that center replicates be added to two-level response surface experiments as a check on the presence of pure quadratic terms and in order to obtain a pure error estimate of σ^2 . The effect of center replicates on the AWEV criterion depends on how the factorial points have been scaled. When d is small and AWEV is dominated by variance rather than bias, adding a center replicate has little effect. When d is large, however, adding a center replicate can reduce AWEV dramatically. Recall that with the Bayesian model proposed here, observations contribute the most information to inferences made at nearby factor combinations. The effect of a center point on AWEV is substantial only when the factorial points are spread so far apart that they provide little information near the origin.

Adding a center replicate results in only a slight decrease in AWEV for the most efficient choices of scale. Only for larger (and typically inefficient) choices of scale is there a large decrease. A beneficial consequence is that the range of efficient designs includes larger values of d and is wider than when no center points are present. Thus including a center point does provide additional robustness with respect to the prior. Adding additional center replicates has only a slight effect in further reducing AWEV.

It is important to remember that an experimental design must satisfy a number of different criteria, of which AWEV reflects but one. We think that the importance of obtaining a pure error estimate of σ^2 is a compelling reason to include several center replicates. The value of a pure error estimate is its independence of any assumptions about the functional dependence of the response variable on the explanatory variables, a property that is especially important for approximate models such as those used here.

7. DISCUSSION

Our conclusions with respect to scaling two-level factorial experiments can be easily summarized: model robust 2^{k-P} designs can be achieved by choosing the scale of the design slightly wider than the scale of the experimenter's weight function. If bias is feared to be especially severe, the design should be pulled in toward the origin, while if bias is suspected to be minimal, the design should be spread out slightly. Importantly, model robust designs are not

highly sensitive to the assumptions about the extent of bias -- large changes in the severity of bias result in only slight changes in the efficient choice of scale. Our conclusions about choice of scale are similar to those of Box and Draper (1959,1963) who advocated "all bias" designs, in which the design scale is chosen to exactly match the weight function. Our conclusions differ substantially, however, from the implication that scale should be as large as possible which results when it is assumed that an empirical model such as (1.2) is an exact representation of the response function.

We have achieved model robustness by using a Bayesian model to represent represent uncertainty about the nature of the true response function. Experimental design must, necessarily, be based on the experimenter's prior knowledge and we think that the Bayesian model offers a natural vehicle to explicitly state prior beliefs about model adequacy. The questionable advice to choose scale as large as possible can thus be seen as a correst conclusion for the implausible prior belief that no bias is present. We have shown that more realistic priors which include bias lead to more sensible conclusions.

Our results in Section 6 on scaling 2^{k-p} designs, although mathematically exact, should be regarded as a guide to choosing an experimental design rather than a prescription. We would be surprised indeed if two scientists, faced with the same problem, arrived at the same list of important factors, assigned them the

same standardized units of measurement, and gave an identical assessment of the bias associated with using a first degree polynomial approximation over the corresponding region of interest. These elements, all of which have an important influence on the final design, must be supplied by the experimenter. The purpose of the methodology presented here is to help the experimenter understand how his region of interest, his prior assumptions about the extent of bias, the number of factors studied, and the extent of fractionation desired should be reflected in the way he scales the design.

References

- Atkinson, A. C. (1972). Planning experiments to detect inadequate regression models. Biometrika, 59, 275-293.
- Box, G. E. P. (1982). Choice of response surface design and alphabetic optimality. Utilitas Mathematica, 21B, 11-55.
- Box, G. E. P. and Draper, N. R. (1959). A basis for the selection of a response surface design. Journal of the American Statistical Association, 622-653.
- Box, G. E. P. and Draper, N. R. (1963). The choice of a second order rotatable design. Biometrika, 50, 335-352.
- Box, G. E. P. and Draper, N. R. (1975). Robust designs. Biometrika, 62, 347-352.
- Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). Statistics for Experimenters. New York: John Wiley and Sons, Inc.
- Conte, S. D. and de Boor, C. (1980). Elementary Numerical Analysis: An Algorithmic Approach, third edition. New York: McGraw-Hill Book Co.
- Galil, Z. and Kiefer, J. (1977). Comparison of rotatable designs for regression on balls, I (quadratic). Journal of Statistical Planning and Inference, 1, 27-40.
- Galil, Z. and Kiefer, J. (1980). Time- and space-saving computer methods, related to Mitchell's DETMAX, for finding D-optimum designs. Technometrics, 22, 301-313.
- Jones, E. R. and Mitchell, T. J. (1978). Design criteria for detecting model inadequacy. Biometrika, 65, 541-551.

- Kussmaul, K. (1969). Protecting against assuming the wrong degree in polynomial regression. Technometrics, 11, 677-682.
- Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. Journal of the Royal Statistical Society, Series B, 34, 1-41 (with discussion).
- Mitchell, T. J. (1974). Computer construction of 'D-optimal' first-order designs. Technometrics, 16, 211-220.
- Mitchell, T. J. and Bayne, C. K. (1978). D-optimal fractions of three-level factorial designs. Technometrics, 20, 369-380.
- Moler, C. (1981). MATLAB Users' Guide.
- O'Hagan, A. (1978). Curve fitting and optimal design for prediction. Journal of the Royal Statistical Society Series B, 40, 1-41.
- Pesotchinsky, L. (1978). Φ_p -optimal second order designs for symmetric regions. Journal of Statistical Planning and Inference 2, 173-188.
- Steinberg, D. M. (1983). Bayesian models for response surfaces and their implications for experimental design. Unpublished doctoral dissertation, University of Wisconsin-Madison.
- Steinberg, D. M. (1984a). Bayesian models for response surfaces I: The equivalence of several models and their relationship to smoothing splines. Mathematics Research Center Technical Summary Report #2682, University of Wisconsin-Madison.

- Steinberg, D. M. (1984b). Bayesian models for response surfaces II:
Estimating the response surface. Mathematics Research Center
Technical Summary Report #2683, University of Wisconsin-Madison.
- Stigler, S. M. (1971). Optimal experimental design for polynomial
regression. Journal of the American Statistical Association, 66,
311-318.
- Wahba, G. (1978). comment to O'Hagan (1978). Journal of the Royal
Statistical Society, Series B, 40, 35-36.
- Watson, G. N. (1933). Notes on generating functions of
polynomials: (2) Hermite polynomials. Journal of the London
Mathematical Society, 8, 194-199.
- Welch, W. J. (1982). Branch-and-bound search for experimental
designs based on D optimality and other criteria. Technometrics,
24, 41-48.
- Young, A. S. (1977). A Bayesian approach to prediction using
polynomials. Biometrika, 64, 309-317.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2721	2. GOVT ACCESSION NO. A147494	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MODEL ROBUST RESPONSE SURFACE DESIGNS: SCALING TWO-LEVEL FACTORIALS		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
7. AUTHOR(s) David M. Steinberg		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of 610 Walnut Street Madison, Wisconsin 53706		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041 MCS-8210950
11. CONTROLLING OFFICE NAME AND ADDRESS (See Item 18 below)		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE July 1984
		13. NUMBER OF PAGES 34
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		16a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES U. S. Army Research Office P. O. Box 12211 Research Triangle Park North Carolina 27709 National Science Foundation Washington, DC 20550		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Response Surface Design; Factorial Experiments; Design Scale; Model Robust Design; Bayesian Models		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Response surface methodology is a useful way to study the relationship between an experimental response variable and a set of continuous explanatory variables. In designing a response surface study, an experimenter must decide how far apart to set the levels of each factor; i.e., how to scale the design. Two conflicting influences must be considered: (i) if the levels are (cont.)		

ABSTRACT (cont.)

too close together, estimates of the response will have high variance, but (ii) if the levels are too far apart, large bias errors may be introduced. We propose a design criterion based on a Bayesian model that makes explicit assumptions about the possible extent of bias and show that the criterion leads to reasonable choices of scale for 2^{k-p} factorial designs. The choice of scale is found to be insensitive to the prior distributions in the model.